

# Predicting Cation Exchange Capacity for Soil Survey Using Linear Models

C. A. Seybold,\* R. B. Grossman, and T. G. Reinsch

## ABSTRACT

Measuring the cation exchange capacity (CEC) for all horizons of every map unit component in a survey area is very time consuming and costly. The objective of this study was to develop CEC (pH 7  $\text{NH}_4\text{OAc}$ ) prediction models that encompass most soils of the United States. The National Soil Survey Characterization database was used to develop the predictive models using general linear models. Data were stratified into more homogeneous groups based on the organic C content, soil pH, taxonomic family mineralogy class and CEC-activity class, and taxonomic order. Models were developed for each strata or data group. Organic matter and noncarbonate clay contents were the main predictor variables used. Water at  $-1500$  kPa was used in lieu of clay content on four groups. Results indicate that between 43 and 78% of the variation in CEC could be explained for the high organic C data groups; between 53 and 84% could be explained for the mineralogy groups; between 86 and 95% could be explained for the CEC-activity class groups; and between 53 and 86% could be explained for the taxonomic orders. The same predictive model was applicable for Gelisols and Histosols, Inceptisols and Alfisols ( $>0.3\%$  organic C) also shared the same model. In general, the mineralogy/CEC-activity class equations had lower RMSEs than the taxonomic order equations. A decision tree, based on how the data was stratified, guides the selection of which model to use for a soil layer. Validation results indicated that the models, in aggregate, provide a reasonable estimate of CEC for most soils of the United States.

CATION EXCHANGE CAPACITY is the total of the exchangeable cations that a soil can hold at a specified pH. Soil components known to contribute to CEC are clay and organic matter, and to a lesser extent, silt (Martel et al., 1978; Manrique et al., 1991). The exchange sites can be either permanent or pH-dependent. Mineral soils have an exchange capacity that is a combination of permanent and pH-dependent charge sites, while that of organic soils is predominantly pH-dependent. In any given soil, the number of exchange sites is dependent on the soil pH; type, size, and amount of clay; and amount, decomposition state, and source of the organic material (Kamprath and Welch, 1962; Parfitt et al., 1995; Syers et al., 1970; Miller, 1970). The relationship between clay content (% by weight) and CEC can be highly variable because different clay minerals have very different CECs, and the relative proportion of pH-dependant and permanent CEC varies among clay minerals (Miller, 1970). Cation exchange capacity of organic soils increases markedly with increases in pH, and increases with greater degrees of humification (Stevenson, 1994). For these rea-

sons, clay and organic matter differences between soils should be incorporated into any kind of predictive model.

Several researchers have attempted to predict CEC from clay and organic C contents alone, using multiple regression. Results show that greater than 50% of the variation in CEC could be explained by the variation in clay and organic C content for several New Jersey soils (Drake and Motto, 1982), for sandy soils in Florida (Yuan et al., 1967), for some Philippine soils (Sahrawat, 1983), and for four soils in Mexico (Bell and van Keulen, 1995). Only a small improvement was obtained by adding pH to the model for four Mexican soils (Bell and van Keulen, 1995). In B horizons of a toposequence, the amount of fine clay ( $<0.2$   $\mu\text{m}$ ) was shown to explain a larger percent of the variation in CEC than the total clay content (Wilding and Rutledge, 1966). In gleyed subsoil horizons of lowland soils in Quebec, surface area (of the soil) gave a better prediction of CEC than did total clay (Martel et al., 1978). Martel et al. (1978) also showed that the variations in mineralogical composition, although small, were sufficient to explain nearly 50% of the variation in CEC. Similarly, Miller (1970) found that the type of clay alone could explain up to 50% of the variation in CEC. Many of the above predictive models are specific to a region or area and confined to only a few soil types. Our approach is to develop predictive models that provide a comprehensive coverage of soils of the United States.

When using least squares estimates in CEC models, the assumption is made that the compositions of the clay and organic matter are identical from one sample to another and that the soils vary only in the amounts of the components present (Stevenson, 1994). For this reason, regression equations tend to be accurate only within a limited geographic and climatic zone, where the composition of the clay and organic fractions are reasonably homogenous (Helling et al., 1964). When soils of diverse genesis are included in the analyses and little or no attempt is made to control for variables such as mineralogical composition, soil properties become less predictive (Syers et al., 1970). When soils are grouped by similarities in origin or properties, accuracy of predictive models (in general) has been shown to improve (Pachepsky and Rawls, 1999). Drake and Motto (1982) grouped soils by taxonomic order or province, which proved superior in defining groups for predicting CEC. Similarly, Asadu and Akamigbo (1990) predicted CEC from organic matter and clay content grouped by taxonomic order (Inceptisols, Alfisols, Ultisols, and Oxisols). They indicated that partitioning the data by taxonomic order resulted in regression equations that were significantly distinct from each other, as the groupings tended to reduce the variability in soil properties. The U.S. Soil

USDA-NRCS, National Soil Survey Center, 100 Centennial Mall North, Federal Building, Room 152, Lincoln, NE 68508. Received 20 Jan. 2004. Pedology. \*Corresponding author (cathy.seybold@nssc.nrcs.usda.gov).

Published in Soil Sci. Soc. Am. J. 69:856–863 (2005).  
doi:10.2136/sssaj2004.0026

© Soil Science Society of America  
677 S. Segoe Rd., Madison, WI 53711 USA

**Abbreviations:** CEC, cation exchange capacity; NASIS, National Soil Information System; OC, organic carbon.

Taxonomy system (Soil Survey Staff, 1999) also classifies soil by mineralogical composition at the family level, which may be useful in partitioning soils to improve both accuracy and reliability of predictive models (Pachepsky and Rawls, 1999).

The National Soil Information System (NASIS) is a relational database of the USDA-NRCS that is used to manage soil survey data. Map unit attributes are stored and maintained in NASIS. Properties in NASIS are estimated, either from direct observation in the field, use of predictive models, or laboratory measurement. Cation exchange capacity values (based on the buffered ammonium acetate method at pH 7) should be populated for most horizons of map unit components with a pH of  $\geq 5.5$  (Soil Survey Staff, 2002). For soil layers with a pH of  $< 5.5$ , the effective cation exchange capacity is populated. In many cases, CEC values (in addition to other properties) are unavailable in the database because of increases in data requirements over time. Also, it is too time-consuming and costly to measure CEC everywhere in a survey. Soil scientists mapping and making entries into NASIS need a reliable method for estimating CEC. The estimation procedure should be comprehensive (encompass most soils of the United States) and be able to predict CEC from accessory or readily available soil properties.

The objectives of this project were to develop CEC prediction models that function comprehensively for the range of U.S. soils. The goal is to use basic soil survey data as input. To improve predictability, the data will be stratified, such as by soil taxonomic order or mineralogy/CEC-activity class. These prediction models will benefit NRCS field soil scientists making entries into NASIS. More importantly, these models should improve the accuracy of estimated CEC data and aid in populating the database, which will benefit all users of soil survey data and their interpretations.

## MATERIALS AND METHODS

Data (pre-1999) from the National Soil Survey characterization database in Lincoln, Nebraska, were used to develop the predictive CEC models. The characterization database contains more than 135 000 horizons with measured CEC data, representing soils from across the continental United States, Hawaii, Alaska, Puerto Rico, and several foreign countries. Relevant data in the database include taxonomic classifications, morphological descriptions, horizon designations, and analytical data such as organic carbon, exchange characteristics, particle-size separates, pH, and water retention characteristics.

### Stratification of Data

The database was partitioned into more homogeneous soil groups to improve the accuracy of CEC estimates. The first division of the data was based on organic C content. An initial query of the database selected records where the pH (in water) was  $\geq 5.5$  and organic C content was  $\leq 8\%$ . The break at 8% organic C is a method break. Organic C by the Walkley-Black method is generally reliable only up to 8% (Soil Survey Staff, 1995). The  $< 8\%$  organic C data group was further subdivided by taxonomic family mineralogy and CEC-activity class. There are 21 taxonomic family mineralogy classes excluding mixed and siliceous (Soil Survey Staff, 1999). Soils with mixed and sili-

ceous mineralogy were grouped by CEC-activity class, where applicable. In Soil Taxonomy, a CEC-activity class is generally assigned to soils with a mixed or siliceous mineralogy at the family level (Soil Survey Staff, 1999). However, there are exceptions to assigning CEC-activity classes. They are not assigned to Histosols or Histels, or to Oxisols, and Alfisols and Ultisols with "kandi" or "kanhap" great groups or subgroups because they would be redundant; all of these would be in a substantive class by definition (Soil Survey Staff, 1999). In addition, CEC-activity classes are also not assigned to soils with sandy, sandy-skeletal, or fragmental particle-size classes. Within the characterization database many pedons did not have taxonomic classifications assigned; therefore, for some mineralogy classes there were little or no data available. Only 12 out of the 21 soil mineralogy classes had enough data from which to develop a model. To ensure that models are developed that cover all soils, the  $< 8\%$  organic C data group was also partitioned by taxonomic order. In a second query, records were selected where the pH (in water) was  $\geq 5.5$  and organic C content was  $\leq 8\%$ , and then subdivided by taxonomic order. There are 12 soil orders in Soil Taxonomy. In total, data were stratified into 12 mineralogy classes and four CEC-activity classes, or 12 soil order groups. The Mollisol and Alfisol orders were the most common and have the largest datasets. Separating out the higher organic matter layers of the Mollisol and Alfisol orders increased the  $R^2$  and lowered the RMSEs of the resulting equations. The organic C break at 0.3% was the point at which the organic C content became an insignificant predictor variable in these two soil orders. The two stratification groupings of the data (by mineralogy/CEC-activity class or soil order) were compared to determine which grouping would explain the most variation in the CEC data, and thus indicate the most homogeneous groups from which models could later be developed. A third grouping was added in the comparison: grouping by horizon designation (e.g., A, Bt, C), which has been used in other studies to stratify the data (Wilding and Rutledge, 1966; Asadu and Akamigbo, 1990). In addition, model RMSEs were compared between the mineralogy/CEC-activity and soil order equations to aid in determining which grouping provides the most accurate estimations. Only the RMSEs on the log transformed scale were compared. Soil layers with high organic C contents ( $> 8\%$ ) were further partitioned into six data groups. Thus, a third query of the database selected records where the pH (in water) was  $\geq 5.5$  and total C was  $> 8\%$ . A plot of the CEC versus pH showed a bimodal distribution with a pH break at 7.0. Therefore, this high organic C data group was subdivided into two groups based on a pH break at 7.0. Then, the  $< 7.0$  pH data group was further subdivided into four groups based on the degree of organic matter decomposition (fibric, hemic, and sapric) and an organic C content break at 14.5%. Decomposition state (fibric, hemic, and sapric) of soils with organic C contents  $\geq 14.5\%$  was indicated by O horizon designation (Oi, Oe, and Oa, respectively). The fourth group consists of an undivided group with an organic C content of  $< 14.5\%$ . The  $> 7.0$  pH data group was subdivided into two groups by an organic C content of 14.5%. The break at 14.5% organic C separates mineral from organic soil material. Soil Survey Staff (1999) defines organic materials as having 12 to 20% organic C depending on the clay content and duration of saturation. An organic C content of 14.5% (approximately 25% organic matter) was chosen as the break between organic and mineral materials for this project. Also, in NASIS, particle size separates (sand, silt, and clay) are generally not populated when organic matter contents are above 25%. For the soil groups with  $< 14.5\%$  organic C, particle-size separates (e.g., clay content) can be used as predictive variables.

### Soil Properties

Variables used in predicting CEC were pH in water, pH in 0.01 *M* CaCl<sub>2</sub>, total clay and total silt (pipette method), noncarbonate clay, organic C (acid-dichromate digestion), total C (dry combustion), and -1500-kPa water (pressure-membrane extraction using sieved samples). Cation exchange capacity was determined by NH<sub>4</sub>OAc at pH 7. All methods are described by the Soil Survey Staff (1996). All determinations were on air-dried (30–35°C), crushed, and sieved (<2 mm) soil samples. Data are reported on oven-dry basis. Carbonate clay has negligible CEC (Shields and Meyer, 1964). Therefore, percent carbonate clay was subtracted from the percent total clay to get noncarbonate clay. This procedure makes it possible to obtain the noncarbonate clay percentage and removes the disadvantage of the particle-size measurement. A -1500-kPa water to clay ratio of >0.6 has been used to indicate poor dispersion in particle size determinations (Soil Survey Staff, 1995, 1999). Poorly crystalline materials and high organic C contents also tend to increase this ratio. When clay or noncarbonate clay was used as a predictor variable, ratios of >0.6 (-1500-kPa water to clay ratio) were excluded from the data. Also, the -1500-kPa water to clay ratio of >0.6 exclusion was not used for any of the high organic C data groups.

### Model Validation

An independent dataset was used to validate the models in aggregate. One-hundred and fifty pedons were randomly selected from the National Soil Survey characterization database, from years 2000 to 2002. The soils represent pedons from all across the United States, including Alaska and Hawaii. Each horizon of each pedon was run through the appropriate predictive model to estimate CEC (using the decision tree). If the horizon pH (in water) was <5.5, no CEC was estimated. As a result, 793 horizons of estimated CEC values were used in the validation process.

### Statistics

For each data group, CEC was estimated using general linear model procedures in SYSTAT Software (2002). Only data elements that contributed significantly ( $P = 0.05$ ) to predicting CEC were used in the regression equations. Also, only variables that contributed >5% to the overall improvement of the  $R^2$  value were included in the equations. Scatter plots of the residuals versus the fitted values of each model were used to indicate whether there was nonlinearity, unequal variances, and outliers in the data. When a classic horn-shaped pattern was evident in the plot, natural log transformations of the data were performed. The horn-shaped pattern indicates a combination of a poor fit to the subpopulation averages and increasing variability (Ramsey and Schafer, 1997). All outliers, as identified by the studentized residual in SYSTAT Software (2002), were removed from the data groups. Pearson correlations were performed to determine variable collinearity and help in the selection of predictive variables. For some data groups (that had low correlation coefficients between variables, e.g., Vertisols), a forward stepwise regression procedure in SYSTAT Software (2002) was used to help identify additional predictive variables from all possible. However, none identified proved to be useful predictive variables. Variables were then added and subtracted from the general linear model until the best model was found that contained statistically significant, intuitively meaningful predictive variables, and variables that are readily obtainable within NASIS. A dummy-variable regressor (taxonomic order or mineralogy) was used to evaluate model redundancy between predictive equations

with the same variables (Fox, 1997). The post hoc Tukey test (multiple mean comparison procedure) was used for comparison of equation intercepts (Zar, 1999). When intercepts between two equations were not significantly different, then the slope coefficients were compared by checking the significance of the interaction terms (dummy-variable and predictive variable). When redundant equations were indicated (no significant difference between slope coefficients and intercepts), the data groups were combined and a new model was developed. Model validation was evaluated by comparing measured versus predicted CEC values from an independent dataset. Confidence intervals were calculated for the slope and intercept of the least square estimate line. Statistically significant differences were determined using  $P = 0.05$ .

## RESULTS AND DISCUSSION

Initial correlation analyses conducted within each data group indicated that total silt, total clay or noncarbonate clay, organic C or total C, pH in water or CaCl<sub>2</sub>, and -1500-kPa water were the variables most highly correlated with CEC (data not shown). Except for -1500-kPa water, these variables are readily available soil properties in NASIS; gravimetric -1500-kPa water can be obtained indirectly within NASIS. The most highly correlated variable with CEC varied, depending on the data group. In general, -1500-kPa water was the single most highly correlated variable with CEC among the data groups. Total clay or noncarbonate clay and -1500-kPa water were highly correlated with each other ( $r > 0.90$ ) in most of the data groups. Therefore, they would be redundant variables if both were included in a regression model. Clay content is a readily available soil property, and is preferred over -1500-kPa water. However, water content at -1500 kPa is preferred as a predictive variable when poor clay dispersion in the particle size determinations is a problem (and/or noncrystalline clays dominate) for a data group. Field-based clay estimates have been determined to be reliable (Nettleton et al., 1999), which were not available for model development. Poor clay dispersion is a problem in the amorphous, glassy, and isotropic mineralogy class data groups, and the Spodosol and Andisol soil order data groups.

There are six high organic C data groups (OC > 8%) for which prediction models were developed (Table 1). For each equation the  $R^2$ , the standard deviation about the regression line (RMSE), and number of samples used ( $n$ ) are presented. For the high organic C data groups with a pH of  $\leq 7.0$ , total C, pH in CaCl<sub>2</sub>, and noncarbonate clay explained between 43 to 63% of the variation in CEC (Table 1). A variable that may help improve the predictability of CEC for these high organic C groups is the fiber content (Lynn et al., 1974). Fiber content was not explored as a potential predictive variable because it is not a readily available data element in NASIS. For the high organic C data groups with a pH of >7.0, total C and noncarbonate clay explained between 78 and 87% of the variability in CEC. Predictability of CEC was greater for soil horizons with a pH of >7.0. Each of the six predictive models is significantly unique. The six prediction equations allow for the calculation of CEC for



**Table 1.** Cation exchange capacity (CEC) linear models and  $R^2$ , root mean square error (RMSE), and  $n$  values for the high organic carbon (OC) and mineralogy/CEC-activity stratification groups.

Grouping	Linear model†	$R^2$	RMSE	$n$
<b>OC &gt; 8% and pH ≤ 7.0</b>				
Eq. [1]; Oa horizons	$2.12(\text{totalC}) + 9.992(\text{pHCaCl}_2) - 10.684$	0.52	27.85	283
Eq. [2]; Oe horizons	$2.03(\text{totalC}) + 3.396(\text{pHCaCl}_2) - 2.939$	0.63	19.61	286
Eq. [3]; Oi horizons	$1.314(\text{totalC}) + 27.047$	0.43	17.01	300
Eq. [4]; OC ≤ 14.5%	$1.823(\text{totalC}) + 0.398(\text{nclay}) + 15.54$	0.42	10.41	133
<b>OC &gt; 8% and pH &gt; 7.0</b>				
Eq. [5]; OC ≤ 14.5%	$\exp[1.316(\ln \text{totalC}) + 1.063(\ln \text{nclay}) - 3.211]$	0.77	0.476‡	275
Eq. [6]; OC > 14.5%	$4.314(\text{totalC}) - 26.492$	0.78	16.62	30
<b>OC ≤ 8%</b>				
Ferruginous	$2.48(\text{OC}) + 0.128(\text{silt}) + 3.208$	0.80	2.01	121
Amorphic	$\exp[0.182(\ln \text{OC}) + 0.817(\ln \text{w15bar}) + 0.736(\ln \text{pHw}) - 0.608]$	0.84	0.262‡	247
Glassy	$\exp[0.102(\ln \text{OC}) + 1.219(\ln \text{w15bar}) - 0.005]$	0.76	0.495‡	257
Carbonatic	$\exp[0.253(\ln \text{OC}) + 0.828(\ln \text{nclay}) + 0.321]$	0.78	0.348‡	406
Magnesic	$2.38(\text{OC}) + 0.555(\text{nclay}) - 0.219(\text{silt}) + 10.428$	0.59	6.27	80
Parasquic	$\exp[0.13(\ln \text{OC}) + 0.65(\ln \text{nclay}) + 0.340(\ln \text{pHw}) - 0.406]$	0.58	0.325‡	258
Micaceous	$\exp[0.251(\ln \text{OC}) + 0.205(\ln \text{clay}) + 0.538(\text{pHw}) - 1.241]$	0.64	0.464	41
Kaolinitic	$\exp[0.206(\ln \text{OC}) + 0.618(\ln \text{nclay}) + 0.303(\ln \text{silt}) + 0.491(\ln \text{pHw}) - 1.786]$	0.56	0.431‡	1 204
Smectitic	$\exp[0.033(\ln \text{OC}) + 0.861(\ln \text{nclay}) + 0.246]$	0.75	0.186‡	1 803
Illitic	$\exp[0.102(\ln \text{OC}) + 0.596(\ln \text{nclay}) - 1.108(\ln \text{pHw}) + 2.892]$	0.67	0.249‡	249
Vermiculitic	$0.365(\text{nclay}) - 9.724(\text{pHw}) + 90.293$	0.75	8.49	40
Isotc	$\exp[0.163(\ln \text{OC}) + 0.683(\ln \text{w15bar}) + 0.812(\ln \text{pHw}) - 0.299]$	0.78	0.329‡	635
Superactive	$\exp[0.039(\ln \text{OC}) + 0.901(\ln \text{nclay}) + 0.131]$	0.90	0.184‡	12 685
Active	$\exp[0.015(\ln \text{OC}) + 0.987(\ln \text{nclay}) - 0.576]$	0.96	0.133‡	4 580
Semiactive	$\exp[0.02(\ln \text{OC}) + 0.974(\ln \text{nclay}) - 0.927]$	0.94	0.189‡	1 648
Subactive	$\exp[0.009(\ln \text{OC}) + 1.02(\ln \text{nclay}) - 1.675]$	0.91	0.289‡	256

† nclay, Noncarbonate clay; pHCaCl<sub>2</sub>, pH in CaCl<sub>2</sub>; pHw, pH in water; w15bar = −1500-kPa water.

‡ Root mean square error (RMSE) or standard deviation of the mean on the natural log transformed scale.

soil layers with an organic C content of >8% and a soil pH in water of ≥5.5.

Prediction models were developed for 12 of the family mineralogy classes in Soil Taxonomy (Table 1). Organic C, pH in water, noncarbonate clay, and −1500-kPa water explained between 56 and 84% of the variability in CEC of the 12 taxonomic family mineralogy class data groups (Table 1). The CEC for the kaolinitic data group was the most difficult to predict ( $R^2 = 0.55$ ), while the amorphous mineralogy class had the highest predictability ( $R^2 = 0.84$ ). Clay dispersion is indicated to be a major problem in the amorphous, glassy, and isotc mineralogy class groups, and thus, −1500-kPa water was used as a predictive variable instead of percent clay. Silt was a useful CEC predictor variable for the ferruginous, magnesian, and kaolinitic mineralogy class groups. Soil pH in water was a useful predictor variable for the amorphous, parasquic, micaceous, kaolinitic, illitic, vermiculitic, and isotc mineralogy groups. The mineralogy and high organic C equations were determined to all be significantly unique. Either the intercept or one of the slope coefficients was significantly different in equations with the same variables. For the regression models presented in this paper, it is assumed that the taxonomic mineralogy class reflects that of the whole soil profile and not just the mineralogy control section. There are cases where this assumption fails such as in soils that have a lithologic discontinuity.

Four CEC-activity class predictive models were developed (Table 1). The CEC-activity classes are assigned to soil with mixed and siliceous mineralogy. Organic C and noncarbonate clay explained between 90 and 96% of the variation in CEC within the CEC-activity class groups (Table 1). The intercepts of the four CEC-activity class models are significantly different from each other,

making them unique equations. All the predictive equations for the soil groups in Table 1 were determined to be significantly unique. Regression equations were developed for all 12 soil orders (Table 2). Organic C, noncarbonate clay, total silt, −1500-kPa water, and pH in water explained between 55 and 86% of the variation in CEC within the 12 soil orders (Table 2). The CEC values for the Vertisol and Oxisol soil orders were the most difficult to predict, with an  $R^2$  of 0.55 and 0.67, respectively. Spodosols and Entisols had the greatest predictability, with  $R^2$  values of 0.86 and 0.85, respectively. Two models with different variables were developed for the Spodosol order; one using noncarbonate clay and the other using −1500-kPa water content (Table 2). The −1500-kPa water explained 86% of the variability in CEC alone for the Spodosol order, while organic C and noncarbonated clay explained only 71% of the variability. In the Andisol order, −1500-kPa water was also a useful variable because of clay dispersion problems. For the Alfisol and Mollisol soil orders, the  $R^2$  tended to improve when the low organic C horizons (<0.3%) were separated from the high organic C horizons. When organic C content was <0.3%, it became an insignificant predictor variable. Tests for redundancy among the CEC models containing the same predictor variables indicated no significant difference in the intercepts or the slope coefficients between the Alfisol (OC > 0.3%) and Inceptisol equations, and between the Gelisol and Histisol equations. These two pairs of equations are considered redundant. The new models of the combined data groups are shown in Table 2.

In comparison with the previous modeling efforts reported in the literature, lower multiple coefficients of determination ( $R^2$ ) for the taxonomic orders were obtained by Manrique et al. (1991). They found that clay

**Table 2.** Cation exchange capacity (CEC) linear models and  $R^2$ , root mean square error (RMSE), and  $n$  values for the taxonomic order stratification groups.

Grouping	Linear model†	$R^2$	RMSE	$n$
Alfisols				
OC $\leq$ 0.3%	$\exp[0.911(\ln \text{ nclay}) - 0.308]$	0.73	0.381‡	4129
OC > 0.3%	$\exp[0.158(\ln \text{ OC}) + 0.805(\ln \text{ nclay}) + 0.216]_a$	0.72	0.305‡	3206
Andisols	$\exp[0.088(\ln \text{ OC}) + 0.885(\ln \text{ w15bar}) + 0.867(\ln \text{ pHw}) - 0.985]$	0.77	0.384‡	1181
Aridisols	$\exp[0.042(\ln \text{ OC}) + 0.828(\ln \text{ nclay}) + 0.236]$	0.75	0.300‡	4114
Entisols	$\exp[0.078(\ln \text{ OC}) + 0.873(\ln \text{ nclay}) + 0.084]$	0.85	0.350‡	1910
Gelisols	$\exp[0.359(\ln \text{ OC}) + 0.49(\ln \text{ clay}) + 1.05]_b$	0.72	0.509‡	97
Inceptisols	$\exp[0.134(\ln \text{ OC}) + 0.794(\ln \text{ nclay}) + 0.239]_a$	0.71	0.421‡	1921
Mollisols				
OC $\leq$ 0.3%	$\exp[0.932(\ln \text{ nclay}) - 0.174]$	0.79	0.285‡	3284
OC > 0.3%	$\exp[0.113(\ln \text{ OC}) + 0.786(\ln \text{ nclay}) + 0.475]$	0.74	0.203‡	8132
Oxisols	$2.738(\text{OC}) + 0.103(\text{nclay}) + 0.123(\text{silt}) - 2.531$	0.67	2.79	781
Spodosols	$\exp[0.045(\ln \text{ OC}) + 0.798(\ln \text{ nclay}) + 0.029]$	0.71	0.311‡	243
	$\exp[0.999(\ln \text{ w15bar}) + 0.317]$	0.86	0.315‡	636
Ultisols	$\exp[0.184(\ln \text{ OC}) + 0.57(\ln \text{ nclay}) + 0.365(\ln \text{ silt}) - 0.906]$	0.76	0.350‡	499
Vertisols	$\exp[0.059(\ln \text{ OC}) + 0.86(\ln \text{ nclay}) + 0.312]$	0.55	0.213‡	2109
Histosols	$\exp[0.319(\ln \text{ OC}) + 0.497(\ln \text{ nclay}) + 1.075]_b$	0.78	0.358‡	60
Gelisols and Histosols	$\exp[0.346(\ln \text{ OC}) + 0.49(\ln \text{ nclay}) + 1.064]$	0.73	0.207‡	157
Alfisols (OC > 0.3%) and Inceptisols	$\exp[0.141(\ln \text{ OC}) + 0.797(\ln \text{ nclay}) + 0.235]$	0.72	0.125‡	5127

† Equations with the same letters are not significantly different from each other. nclay, Noncarbonate clay; pHw, pH in water; w15bar,  $-1500\text{-kPa}$  water; OC, organic carbon.

‡ Root mean square error (RMSE) or standard deviation of the mean on the natural log transformed scale.

and organic C accounted for up to 67% of the variation in CEC for Alfisols, Inceptisols, Mollisols, and Vertisols, and up to 78% of the variation in CEC for Entisols and Spodosols. The data used to develop these models are from the same database as was used in the present study. In the present study, data with clay dispersion problems were removed and data transformations were conducted, which may explain the differences in the ability to explain variation in CEC (for soil orders) in this study and that of Manrique et al. (1991). Asadu and Akamigbo (1990) also developed CEC prediction models for only four of the soil orders by horizon. They found organic matter and clay content to explain between 42% (in Alfisols) to 80% (in Ultisols) of the variation in CEC for all the A horizons and 23% (in Oxisols) to 67% (in Inceptisols) for all the B horizons. For horizons of Andisols that have andic properties, Nettleton et al. (2001) found that 69% of the variation in CEC could be explained by the organic C content alone.

The range in property values for each predictive variable used in the development of each equation is presented in Table 3. Prediction of CEC for each individual equation is valid only within the property range of the predictive variables used to develop the model (Ramsey and Schafer, 1997). Prediction of CEC and use of the regression equations are limited to the range of properties used in this study, which encompasses most soils of the United States.

### Model Selection

In Tables 1 and 2, there are two sets of equations (taxonomic order or mineralogy/CEC-activity equations) that can be used to predict CEC for soil layers with less than 8% organic C. When more than one prediction equation is available for a particular soil, the most accurate equation should be used (Pachepsky and Rawls, 1999). Three data grouping variables were compared—by taxonomic family mineralogy/CEC-activity class, taxonomic order, and horizon designation. Taxonomic min-

eralogy/CEC-activity variable explained the most variation in CEC ( $r^2 = 0.30$ ) followed by soil order ( $r^2 = 0.21$ ) and then horizon designation ( $r^2 = 0.10$ ). Out of the three grouping variables, taxonomic mineralogy/CEC-activity may provide for the most homogeneous soil groups to improve accuracy of estimating CEC. In support of this, the mean RMSE values of the mineralogy/CEC-activity class equations were significantly lower than the mean RMSE of the soil order regression equations ( $P = 0.036$ ). This may suggest that the mineralogy/CEC-activity class equations, as a group, might be more accurate than the soil order regression equations. Therefore, the mineralogy class/CEC-activity class prediction equations should be used first, then the taxonomic order equations. However, there would be some soils where CEC would not be predicted if only the mineralogy/CEC-activity models were used in Table 1. If an equation does not exist for a mineralogy class (e.g., ferritic), then the taxonomic order equations are recommended to be used. This recommendation along with the pH and organic C data breaks for the remaining equations are presented in Fig. 1 as a decision tree. The tree is a guide for using the regression equations. For a given soil layer, if the soil pH is  $\geq 5.5$ , then CEC is estimated. If the organic C content is  $> 8\%$ , then the tree goes through the data breaks for using the high organic C predictive models. If the organic C content is  $\leq 8\%$ , then the tree, first, uses the mineralogy/CEC-activity class equations, and then the soil order equations. As soon as the criteria match for a horizon, that particular predictive model is used to estimate CEC. The decision tree provides a CEC estimate for every soil layer with a pH of  $\geq 5.5$ , given that the pH, organic C content, and soil classification are known.

### Model Validation

A plot of the measured versus predicted CEC values for 793 horizons from 150 pedons is shown in Fig. 2. Most of the pedons have mixed or siliceous mineralogy

**Table 3.** Range in properties of variables used to predict cation exchange capacity (CEC) for the high total carbon (TC) and the family mineralogy/CEC-activity class data groups, and taxonomic order data groups with soil pH values of  $\geq 5.5$  and organic carbon (OC) contents of  $\leq 8\%$ .

Grouping	Organic C			Noncarbonate clay			Soil pH (1:1 water)			Silt or -1500-kPa water		
	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD
	%						%					
Oa horizon	11.6–71 <sup>†</sup>	30.0	13.7				2.3–6.8 <sup>‡</sup>	4.2	1.1			
Oe horizon	11.6–63 <sup>†</sup>	32.5	13.2				2.5–6.8 <sup>‡</sup>	4.3	0.9			
Oi horizon	11.6–71 <sup>†</sup>	37.4	11.4									
Eq. [4]	8.1–14.5 <sup>†</sup>	10.4	1.9	0.0–89.0	17.7	15.6						
Eq. [5]	8.1–14.4 <sup>†</sup>	9.7	1.6	0.1–75.5	18.6	13.0						
Eq. [6]	14.6–41 <sup>†</sup>	21.5	6.6									
Ferruginous	0.0–7.8	1.26	1.74							8.3–63.6 <sup>§</sup>	28.8	13.7
Amorphous	0.04–7.9	3.15	2.16				4.0–9.2	6.09	1.03	1.8–56.7 <sup>¶</sup>	21.8	10.9
Glassy	0.0–7.8	1.1	1.6							1.4–27.0 <sup>¶</sup>	8.1	1.1
Carbonatic	0.01–7.8	1.09	1.22	2.2–56.6	20.0	9.99						
Micaceous	0.06–5.8	1.12	1.37	0.3–54.1	14.6	11.4	5.5–8.1	6.3	0.8			
Magnesian	0.16–6.0	1.7	1.4	8.8–55.6	29.8	11.5				8.2–57.1 <sup>§</sup>	31.7	10.7
Parasquic	0.02–5.9	0.83	1.03	6.1–81.8	40.1	16.9	3.6–6.7	5.2	0.6			
Kaolinitic	0.01–7.73	0.84	1.04	2.8–94.7	49.5	21.5	3.5–8.2	5.3	0.8	1.7–78.2 <sup>§</sup>	20.3	12.3
Smectitic	0.01–7.1	0.8	0.8	6.8–82.3	39.2	12.8						
Illitic	0.01–4.35	0.62	0.7	7.7–79.5	38.9	15.4	5.5–10.3	7.4	0.95			
Vermiculitic				4.8–59.6	30.8	17.4	6.0–8.6	7.2	0.9			
Isotie	0.02–7.9	1.5	1.7				3.9–8.4	5.5	0.75	0.7–41.0 <sup>¶</sup>	9.6	6.3
Superactive	0.01–8.0	0.73	0.81	0.1–86.9	22.1	11.2						
Active	0.01–5.6	0.42	0.47	1.6–91.5	26.7	14.9						
Semiactive	0.01–3.7	0.29	0.28	0.1–94.6	28.6	18.1						
Subactive	0.01–2.9	0.27	0.34	0.1–94.9	29.7	20.5						
Alfisols												
OC $\leq 0.3\%$				0.1–94.6	24.8	15.0						
OC $> 0.3\%$				2.0–94.4	28.4	15.6						
Andisols	0.01–8.0	1.54	1.70				5.5–9.2	6.2	0.60	1.0–60.3 <sup>¶</sup>	13.5	9.2
Aridisols	0.01–5.1	0.47	0.45	0.1–84.8	22.3	12.8						
Entisols	0.01–5.5	0.54	0.58	0.1–89.0	20.7	15.6						
Gelisols	0.03–7.07	1.31	1.48	2.3–94.9 <sup>#</sup>	34.1	30.2						
Inceptisols	0.01–7.3	0.60	0.75	0.1–92.5	25.0	16.8						
Mollisols												
OC $\leq 0.3\%$				1.5–78.0	24.6	11.6						
OC $> 0.3\%$				3.2–92.8	30.4	12.0						
Oxisols	0.03–6.1	1.07	1.17	6.1–94.7	58.2	18.2				3.3–58.0 <sup>§</sup>	21.4	11.7
Spodosols												
Eq. [1]	0.01–4.2	0.18	0.42	0.8–46.6	9.2	6.1						
Eq. [2]										0.4–43.9 <sup>¶</sup>	5.2	4.7
Ultisols	0.02–5.9	0.70	0.83	2.0–92.2	29.9	19.0				2.4–79.2 <sup>§</sup>	30.4	18.1
Vertisols	0.01–5.8	0.79	0.75	15.3–87	51.4	11.6						
Histosols	0.11–7.9	1.91	2.17	1.3–95	21.8	15.6						

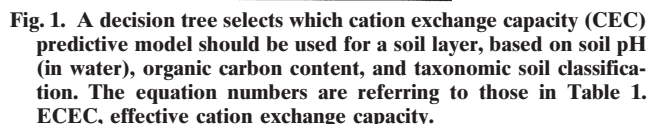
<sup>†</sup> Percent total C.<sup>‡</sup> pH in CaCl<sub>2</sub>.<sup>§</sup> Percent silt.<sup>¶</sup> -1500-kPa water.<sup>#</sup> Percent clay.

classifications with an assigned CEC-activity class. The breakdown of the soil classifications are: 2% of the soil layers had  $>8\%$  organic C, 29% had a taxonomic mineralogy class other than mixed or siliceous, 44% had a CEC-activity class, and 25% were estimated based on the taxonomic soil order. Of the mineralogy classes, smectitic was the most common. Superactive was the most common CEC-activity class. All developed prediction models were used except the vermiculitic mineralogy class and Oxisol soil order equations. The coefficient of determination ( $r^2$ ) was 0.87 and RMSE was 6.176. When the high organic C horizons (OC  $> 8\%$ ) were excluded, the RMSE decreased to 4.494, while the  $r^2$  remained the same. This indicates that the high organic C predictive models are less reliable in predicting CEC. The 95% confidence intervals about the slope were 0.952 to 1.004, which includes unity; there is no significant difference between the slope and unity. The 95% confidence intervals about the intercept were  $-0.001$  to 1.292, which does include zero, which indicates the inter-

cept is not significantly different from zero. This suggests that the regression models, in aggregate, can provide a reasonable estimate of CEC with decreasing reliability at greater organic C contents. Since these models are based on Soil Taxonomy, it is critical that the soils are classified correctly, especially the mineralogy and CEC-activity classes. In soil survey, soil scientists will generally estimate the soil classification based on experience and knowledge of the soils in the area, the morphology, and maybe some lab data.

## CONCLUSIONS

There are 12 family mineralogy class and four CEC-activity class equations, and 10 taxonomic order equations. Six equations were developed for high organic C content soil layers (grouped by pH and organic C content). In total, 28 unique predictions equations were developed. Dominate variables used in the development of the models were organic C content, clay and silt con-



A scatter plot comparing Measured CEC (cmol<sub>(+)</sub> kg<sup>-1</sup>) on the y-axis to Predicted CEC (cmol<sub>(+)</sub> kg<sup>-1</sup>) on the x-axis. Both axes range from 0 to 200. A solid diagonal line represents the 1:1 relationship. Data points are categorized by soil type: Mineral (solid circles) and > 8% Organic C (solid triangles). Mineral soils are clustered at lower CEC values (0-50), while > 8% Organic C soils show a wider range (40-185). Most points fall above the 1:1 line, indicating that the model tends to predict a higher CEC than what was measured.

**Fig. 2. A scatter plot of the measured vs. predicted cation exchange capacity (CEC) of 793 horizons from 150 pedons from across the United States.**

## REFERENCES

- Asadu, C.L.A., and F.O.R. Akamigbo. 1990. Relative contribution of organic matter and clay fractions to cation exchange capacity of soils in southern Nigeria. *Samaru J. Agric. Res.* 7:17–23.
- Bell, M.A., and J. van Keulen. 1995. Soil pedotransfer functions for four Mexican soils. *Soil Sci. Soc. Am. J.* 59:865–871.
- Drake, E.H., and H.L. Motto. 1982. An analysis of the effect of clay and organic matter content on the cation exchange capacity of New Jersey soils. *Soil Sci.* 133:281–288.
- Fox, J. 1997. *Applied regression analysis, linear models, and related methods.* Sage Publ., Thousand Oaks, CA.
- Helling, C.S., G. Chesters, and R.B. Corey. 1964. Contribution of organic matter and clay to soil cation-exchange capacity as affected by the pH of the saturating solution. *Soil Sci. Soc. Am. Proc.* 28:517–520.
- Kamprath, E.J., and C.D. Welch. 1962. Retention and cation-exchange properties of organic matter in coastal plain soils. *Soil Sci. Soc. Am. Proc.* 26:263–265.
- Lynn, W.C., W.E. McKinzie, and R.B. Grossman. 1974. Field laboratory tests for characterization of Histosols. p. 11–20. *In* A.R. Adenah, S.W. Boul, D.E. Hill, and H.H. Bailey (ed.) *Histosols: Their characterization, classification and use.* SSSA Spec. Publ. 6. SSSA, Madison, WI.
- Manrique, L.A., C.A. Jones, and P.T. Dyke. 1991. Predicting cation-exchange capacity from soil physical and chemical properties. *Soil Sci. Soc. Am. J.* 55:787–794.
- Martel, Y.A., C.R. De Kimpe, and M.R. Laverdiere. 1978. Cation-exchange capacity of clay-rich soils in relation to organic matter, mineral composition, and surface area. *Soil Sci. Soc. Am. J.* 42:764–767.
- Miller, W.F. 1970. Inter-regional predictability of cation-exchange capacity by multiple regression. *Plant Soil* 33:721–725.
- Nettleton, W.D., S.H. Brownfield, E.C. Benham, R. Burt, K. Hipple, C.L. McGrath, and H.R. Sinclair, Jr. 2001. Predictive models for selected chemical properties of Andisols. *Soil Survey Horizons* 42:99–111.
- Nettleton, W.D., S.H. Brownfield, R. Burt, E.C. Benham, S.L. Baird, K. Hipple, C.L. McGrath, and H.R. Sinclair. 1999. Reliability of Andisol field texture clay estimates. *Soil Survey Horizons* 40:36–49.
- Pachepsky, Y.A., and W.J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Sci. Soc. Am. J.* 63:1748–1757.
- Parfitt, R.L., D.J. Giltrap, and J.S. Whitton. 1995. Contribution of organic matter and clay minerals to the cation exchange capacity of soils. *Commun. Soil Sci. Plant Anal.* 26:1343–1355.
- Ramsey, F.L., and D.W. Schafer. 1997. *The statistical sleuth: A course in methods of data analysis.* Wadsworth Publ., Belmont, CA.
- Sahrawat, K.L. 1983. An analysis of the contribution of organic matter and clay to cation exchange capacity of some Philippine soils. *Commun. Soil Sci. Plant Anal.* 14:803–809.
- Shields, L.G., and M.W. Meyer. 1964. Carbonate clay: Measurement and relationship to clay distribution and cation-exchange capacity. *Soil Sci. Soc. Am. Proc.* 28:416–419.
- Soil Survey Staff. 1995. *Soil survey laboratory information manual.* Soil Survey Investigations Rep. 45. Version 1.0. U.S. Gov. Print. Office. Washington, DC.



- Soil Survey Staff. 1996. Soil survey laboratory methods manual. Soil Survey Investigations Rep. 42. Version 3.0. U.S. Gov. Print. Office, Washington, DC.
- Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd ed. Agric. Handb. 436. U.S. Gov. Print. Office, Washington, DC.
- Soil Survey Staff. 2002. National soil survey handbook [Online]. Available at <http://soils.usda.gov/technical/handbook/> (verified 21 Jan. 2005). USDA-NRCS, Lincoln, NE.
- Stevenson, F.J. 1994. Humus chemistry: Genesis, composition, reactions. 2nd ed. John Wiley & Sons, New York.
- Syers, J.K., A.S. Campbell, and T.W. Walker. 1970. Contribution of organic carbon and clay to cation exchange capacity in a chronosequence of sandy soils. *Plant Soil* 33:104–112.
- SYSTAT Software. 2002. SYSTAT for windows. Version 10.2. SYSTAT Software, Richmond, CA.
- Wilding, L.P., and E.M. Rutledge. 1966. Cation-exchange capacity as a function of organic matter, total clay, and various clay fractions in a soil toposequence. *Soil Sci. Soc. Am. Proc.* 30:782–785.
- Yuan, T.L., N. Gammon, Jr., and R.G. Leighty. 1967. Relative contribution of organic and clay fractions to cation-exchange capacity of sandy soils from several soil groups. *Soil Sci.* 104:123–128.
- Zar, J.D. 1999. Biostatistical analysis. 4th ed. Prentice Hall, Upper Saddle River, NJ.